

Original Paper

doi [10.15826/recon.2020.6.3.011](https://doi.org/10.15826/recon.2020.6.3.011)

JEL I18, P46, R11



Socio-economic factors in the spread of SARS-COV-2 across Russian regions

E.V. Sinitsyn¹ ✉, A.V. Tolmachev^{1,2}, A.S. Ovchinnikov¹¹ Ural Federal University, Ekaterinburg, Russia; sinitsyn_ev@mail.ru² Datatel-Ural, Ltd., Ekaterinburg, Russia**ABSTRACT**

Relevance. The worldwide spread of a new infection SARS-CoV-2 makes relevant the analysis of the socio-economic factors that make modern civilization vulnerable to previously unknown diseases. In this regard, the development of mathematical models describing the spread of pandemics like COVID-19 and the identification of socio-economic factors affecting the epidemiological situation in regions is an important research task. **Research objective.** This study seeks to develop a mathematical model describing the spread of COVID-19, thus enabling the analysis of the main characteristics of the spread of the disease and assessment of the impact of various socio-economic factors. **Data and methods.** The study relies on the official statistical data on the pandemic presented on coronavirus sites in Russia and other countries, Yandex DataLens dataset service, as well as data from the Federal State Statistics Service. The data were analyzed by using a correlation analysis of COVID-19 incidence parameters and socio-economic characteristics of regions; multivariate regression – to determine the parameters of the probabilistic mathematical model of the spread of the pandemic proposed by the authors; clustering – to group the regions with similar incidence characteristics and exclude the regions with abnormal parameters from the analysis. **Results.** A mathematical model of the spread of the COVID-19 pandemic is proposed. The parameters of this model are determined on the basis of official statistics on morbidity, in particular the frequency (probability) of infections, the reliability of the disease detection, the probability density of the disease duration, and its average value. Based on the specificity of COVID-19, Russia regions are clustered according to disease-related characteristics. For clusters that include regions with typical disease-related characteristics, a correlation analysis of the relationship between the number of cases and the rate of infection (with the socio-economic characteristics of the region is carried out. The most significant factors associated with the parameters of the pandemic are identified. **Conclusions.** The proposed mathematical model of the pandemic and the established correlations between the parameters of the epidemiological situation and the socio-economic characteristics of the regions can be used to make informed decisions regarding the key risk factors and their impact on the course of the pandemic.

KEYWORDS

COVID-19, mathematical model of the epidemic spread, probability of infection and disease detection, density of disease duration probability, correlation analysis, socio-economic determinants of the COVID-19 pandemic

FOR CITATION

Sinitsyn, E.V., Tolmachev, A.V., & Ovchinnikov, A.S. (2020) Socio-economic factors in the spread of SARS-COV-2 across Russian regions. *R-economy*, 6(3), 129–145. doi: 10.15826/recon.2020.6.3.011

Социально-экономические факторы распространения SARS-COV-2 по регионам России

Е.В. Синицын¹ ✉, А.В. Толмачёв^{1,2}, А.С. Овчинников¹¹ Уральский федеральный университет, Екатеринбург, Россия; sinitsyn_ev@mail.ru² Датател-Урал, Екатеринбург, Россия**АННОТАЦИЯ**

Актуальность. Распространение по всему миру новой инфекции SARS-CoV-2 делает актуальным анализ социально-экономических факторов, которые делают современную цивилизацию уязвимой перед неизвестными ранее заболеваниями. В связи с этим разработка математических моделей, описывающих распространение пандемий типа COVID-19, и выявление социально-экономических факторов, влияющих на эпидемиологическую ситуацию в регионах, является важной исследовательской задачей. **Цель исследования.** Это исследование направлено на разработку математической модели, описывающей распространение COVID-19, что позволит проанализировать основные характеристики распространения заболевания

КЛЮЧЕВЫЕ СЛОВА

COVID-19, математическая модель распространения эпидемии, вероятность заражения и выявления заболевания, плотность вероятности продолжительности заболевания, корреляционный анализ, социально-экономические детерминанты пандемии COVID-19

© Sinitsyn, E.V., Tolmachev, A.V., Ovchinnikov, A.S., 2020

и оценить влияние различных социально-экономических факторов. **Данные и методы.** В основе исследования лежат официальные статистические данные о пандемии, представленные на российских и зарубежных сайтах, сервисе данных Yandex DataLens, а также сайте Росстата. Данные были проанализированы с помощью корреляционного анализа параметров заболеваемости COVID-19 и социально-экономических характеристик регионов; многомерная регрессия была использована для определения параметров вероятностной математической модели распространения пандемии, предложенной авторами; кластеризация позволила сгруппировать регионы со схожими характеристиками заболеваемости и исключить из анализа регионы с аномальными параметрами. **Результаты.** Предложена математическая модель распространения пандемии COVID-19. Параметры этой модели определяются на основе официальной статистики заболеваемости, в частности частоты (вероятности) инфекций, надежности выявления болезни, плотности вероятности продолжительности болезни и ее среднего значения. Исходя из специфики COVID-19, регионы России сгруппированы по характеристикам, связанным с заболеванием. Для кластеров, которые включают регионы с типичными характеристиками, связанными с заболеванием, проводится корреляционный анализ взаимосвязи между количеством случаев и уровнем инфицирования (R_t) с социально-экономическими характеристиками региона. Выявлены наиболее значимые факторы, связанные с параметрами пандемии. **Выводы.** Предложенная математическая модель пандемии и установленные корреляции между параметрами эпидемиологической ситуации и социально-экономическими характеристиками регионов могут быть использованы для принятия обоснованных решений относительно ключевых факторов риска и их влияния на течение пандемии.

ДЛЯ ЦИТИРОВАНИЯ

Sinitsyn, E.V., Tolmachev, A.V., & Ovchinnikov, A.S. (2020) Socio-economic factors in the spread of SARS-COV-2 across Russian regions. *R-economy*, 6(3), 129–145. doi: 10.15826/recon.2020.6.3.011

Introduction

The rapid worldwide spread of the new SARS-CoV-2 infection has raised some important questions about the socio-economic problems that make modern civilization vulnerable to new, previously unknown diseases. To prevent situations like the COVID-19 pandemic in the future, its development should be analyzed and its patterns should be identified, in particular, to assess the influence of various socio-economic factors on its parameters such as the disease spread rate (number of cases, growth rate) and its peculiarities (disease duration and mortality). The results of this analysis can be used to build models of disease spread and make decisions concerning disease prevention.

The problems of modeling of global disease spread have been discussed for a long time. A mathematical model of influenza spread developed by Rvachev (1971) and the models of nonlinear population waves developed by Svirezhev (1987) can serve as an example. The issues of construction of probabilistic epidemics models were considered by Whittle (1955), who formulated the equations for the distribution of the number of individuals infected during an epidemic. This distribution turns from unimodal to bimodal with a decrease in the ratio of the intensity of isolation of infected individuals to the intensity of infection of healthy individuals to some critical value (Bailey,

1975). In more recent works (Barlow & Weinstein, 2020), the so-called SIR (Susceptible – Infected – Recovered) epidemics models were considered and differential equations for the number of infected individuals were formulated. Studies that focused on the spread of coronavirus proposed extensions of the SIR model by using nonlinear differential equations describing the dynamics of various groups of participants in the epidemic (e.g. Ndairou et al., 2020; Abdo et al., 2020). A large number of works are concerned with the algorithms for epidemic spread prediction, based on both probabilistic models (e.g. Zhang *et al.*, 2020) and more traditional approaches based on ARIMA, where epidemiological indicators are linked to a number of social and demographic characteristics of countries affected by the COVID-19 pandemic (e.g. Chakraborty & Ghosh, 2020). Finally, some studies apply various theoretical models for calculation of the reproduction index – the expected number of secondary cases caused in a fully susceptible population by a typical infected person (Driessche & Watmough, 2020; Ndairou et al., 2020).

Nevertheless, many factors relevant to decision-making in a pandemic, influencing the spread of the disease and leading to differences in the way the pandemic progressed in different regions remain undetected. In these conditions, to borrow other's countries socio-economic strat-

egies is not the optimal way as it can result in negative consequences for the country's people and economy.

To address the above research gap, this study intends to meet the following objectives:

- to compare the statistical data on the progress of the COVID-19 pandemic in Russian regions;

- to identify socio-economic factors that determined the main characteristics of the disease spread;

- to construct a probabilistic mathematical model describing the progress of the pandemic that would enable us to forecast the course of pandemic and assess the reliability of such forecasts.

The study relies on the materials retrieved from the Yandex DataLens¹ service, the official website on coronavirus in the Russian Federation² as well as the Rosstat data for Russian regions (2018)³.

1. Theoretical foundations of the analysis of the COVID-19 pandemic

As already noted in the literature described above, a large number of problems in a variety of fields can be reduced to random-walk processes of the subjects under study along the nodes of a directed graph. Moreover, each node corresponds to a particular group (cluster) in which the subjects under study are united in accordance with their inherent characteristics. This can be, for example, customers choosing the products of one of the competing manufacturers (Sinitsyn et al., 2011), vehicles moving along city streets between the intersections where they congregate (Tolmachev et al., 2019), and even students belonging to different performance groups in massive open online courses.

In the model of the COVID-19 pandemic considered herein, the population is grouped by health status as follows:

I – infected individuals in whom the disease has not been identified: those who are ill without clinical symptoms, who did not apply to medical institutions or were mistakenly diagnosed with the absence of the disease;

H – healthy and susceptible to disease individuals;

¹ <https://datalens.yandex.ru> (date of access: August 10, 2020)

² <https://стопкоронавирус.рф> (date of access: August 10, 2020)

³ Regions of Russia. Socio-Economic Indicators (2018). P32. Rosstat. Moscow, 1162.

A – detected active cases of the disease;

R – recovered;

D – deaths;

U – disease-resistant (disease-resistant individuals: the cases of infected persons who were undetected and then recovered as well as vaccinated persons).

Hereinafter, the size of each group will be denoted by letter N with the corresponding index $N_I \dots N_U$. The relative share of each group in the population of region N will be denoted as n with an index denoting the group: $n_I \dots n_U$.

$$N_I + N_H + \dots + N_U = N$$

or

$$n_I + n_H + \dots + n_U = 1 \quad (1)$$

The graph corresponding to the model used herein is shown in Figure 1.

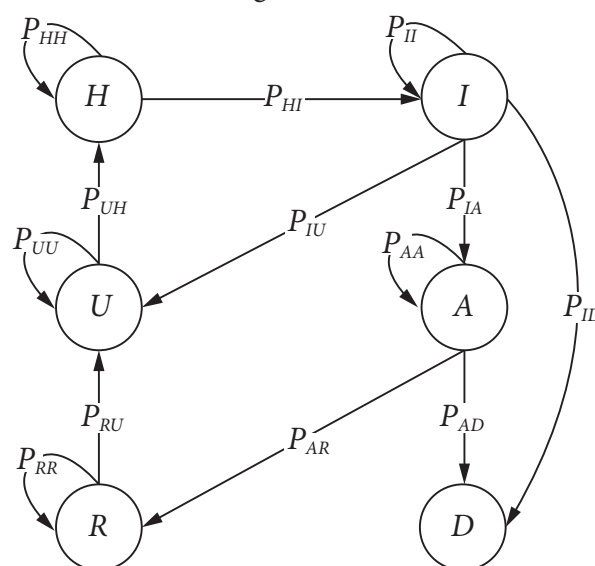


Figure 1. The graph of possible transitions between groups of participants in the epidemic. The indicators on the arrows are transition probabilities

The following assumptions arising from the information about the course of the coronavirus-induced disease contained on the sites on coronavirus in the Russian Federation⁴ and other countries⁵ were made when constructing the graph:

1. The recovery is accompanied by persistent immunity (the probability of re-infection is formally taken into account by the probability of P_{UH} , hereinafter the authors will assume $P_{UH} = 0$).

⁴ <https://стопкоронавирус.рф> (date of access: August 10, 2020)

⁵ <https://www.worldometers.info/coronavirus/> (date of access: August 10, 2020)

2. All the recovered persons recover completely, that is, there is no return of those who have recovered to the group of active cases (treatment is not completed or the conclusion about the recovery is wrong).

3. Long-term consequences leading to death after a fixed recovery (transitions between groups R and D) are not taken into account⁶.

4. There are no vaccinations, that is, transitions from the H group directly to the U group are impossible.

If necessary, the above restrictions can be removed and the model can be expanded, but this will lead to computational complexity.

The transitions shown in Figure 1 can be described by using a 6×6 matrix \hat{T} , which, taking into account the above remarks, has the following form:

$$\hat{T} = \begin{pmatrix} I & H & A & R & D & U \\ P_{II} & P_{HI} & 0 & 0 & 0 & 0 \\ 0 & P_{HH} & 0 & 0 & 0 & 0 \\ P_{IA} & 0 & P_{AA} & 0 & 0 & 0 \\ 0 & 0 & P_{AR} & P_{RR} & 0 & 0 \\ P_{ID} & 0 & P_{AD} & 0 & P_{DD} & 0 \\ P_{IU} & 0 & 0 & P_{RU} & 0 & P_{UU} \end{pmatrix} \begin{matrix} I \\ H \\ A \\ R \\ D \\ U \end{matrix} \quad (2)$$

The columns of matrix (2) contain the probabilities of transitions from the group indicated in the top row to the group presented in the rightmost column, which implies that the sum of all elements of each column is equal to 1. Such matrices are usually called stochastic (Leskovec et al., 2020).

Using matrix (2), it is easy to formulate the equation for probability $P(N_I, N_H, \dots, N_U; y) \equiv P(\vec{X}; t) \equiv P$ that groups $I - U$ at the time unit t will contain N_I, N_H, \dots, N_U individuals. Here $\vec{X} = \{N_I, \dots, N_U\}$ is the vector with components equal to the sizes of the corresponding groups. Performing simple calculations by analogy with the approach described, for example, by Feller (1964) when deriving the equations for the processes of birth and death, one can get the following:

$$\frac{\partial P}{\partial t} = P\{T_{kk}N_k - N\} + T_{k \neq j}(N_k + 1)P(\dots, N_k + 1, \dots, N_j - 1, \dots; t). \quad (3)$$

⁶ In reality, such situations may arise, for example, due to the exacerbation of chronic diseases provoked by the COVID-19 disease, however, there are no reliable statistics on such facts.

It is implied that repeated indices are summed.

Obviously, (3) is an analog of the Kolmogorov equation for the considered model. Using (3), various aspects connected with probability $P(N_I, N_H, \dots, N_U; t)$ can be studied; in particular, it is possible to determine confidence intervals for all projections of morbidity, mortality, and recovery rates as well as to analyze possible epidemic scenarios and assess their probabilities. However, it is more appropriate to discuss these results in a special work concerned with the mathematical aspects of the model under consideration. For the purposes of further analysis, this paper is restricted to the derivation on the basis of (3) of the equations for the mathematical expectations of values n_I, n_H, \dots, n_U ⁷. In discrete form⁸ after a series of transformations, one can obtain:

$$n_H(t+1) = n_H(t) - P_{HI}n_H(t) \quad (4a)$$

$$n_I(t+1) = n_I(t) - P_{IA}n_I(t) - P_{IU}n_I(t) + P_{HI}n_H(t) \quad (4b)$$

$$n_A(t+1) = n_A(t) - P_{AD}n_A(t) - P_{AR}n_A(t) + P_{IA}n_I(t) \quad (4c)$$

$$n_D(t+1) = n_D(t) + P_{AD}n_A(t) \quad (4d)$$

$$n_U(t+1) = n_U(t) + P_{RU}n_R(t) + P_{IU}n_I(t) \quad (4e)$$

$$n_R(t+1) = n_R(t) + P_{AR}n_A(t) - P_{RU}n_R(t) \quad (4f)$$

When deriving equations (4), the above-mentioned feature of matrix (2) was used (the sum of the elements of each column is equal to 1). It was additionally assumed that $P_{ID} = 0$. The reason for this is that all severe cases ending in death require medical intervention and are recorded as active before death (transition from group I to group A). Those who are asymptomatic or mildly symptomatic and do not seek medical help and accordingly, are unlikely to die may well remain undetected.

Let us make one more well-grounded assumption: let us exclude the probability of delayed formation of immunity (thus the subjects from group R (recovered) will go to group U (disease resistant) immediately after their recovery, without participating in group R for the time required for the final formation of immunity). This is con-

⁷ This approach, with the exception of the number of analyzed participants in the epidemic, is similar to the SIR model described above.

⁸ Equations (4) can be written in differential form similarly to (3). To do this, it is required to transfer $n_H(t)$ to the left side of the equation and replace,

$$[n_H(t+1) - n_H(t)] \rightarrow \frac{\partial n_H}{\partial t},$$

etc. for all other equations.

sistent with the fact that $P_{RU} = 1$, while $P_{RR} = 0$. It is obvious that in this case, group U actually absorbs group R , and equations (4e) and (4f) can be replaced by one:

$$n_U(t+1) = n_U(t) + P_{AR}n_A(t) + P_{IU}n_I(t) \quad (4g)$$

It should also be noted that the selection of two different groups of infected people (I) and active cases (A) is based on the assumption that not all cases of COVID-19 are detected. Therefore, the probability of transition between these groups is different from 1. The available statistical information does not have enough data to determine the size of group I . As noted in the monograph by Svirezhev (1987), the probability P_{IA} can be represented as:

$$P_{IA} = P_\alpha \left(d + \frac{T}{N} \right). \quad (5)$$

Here d is the likelihood of an infected person self-referring to a doctor per time unit, T is the number of tests carried out per time unit to detect COVID-19⁹, and finally, P_α is the reliability of disease detection – the probability of a type I error in diagnosis (the disease goes detected).

An attempt can be made to determine parameters (5) using equation (4c) and data on the number of active cases and the number of tests. For example, Figure 2 shows the proportion of detected COVID-19 cases in the tests performed.

⁹ In (5), the number of tests is divided by with the population size.

The initial phase of the pandemic is clearly visible (until about mid-May 2020). In this phase, an exponential increase in the number of cases and positive results of tests performed was observed¹⁰. Its subsequent stabilization followed by a decline to a certain constant value is a characteristic form of such graphs for most countries. For example, Figure 2 shows the graphs for Russia and Germany.

Assuming that the concentration of cases determined for the tested groups can be extended to the entire population of Russia, a simple calculation shows that the number of infected people in the country should be about 4.25 million people, which significantly exceeds the official data of about 860 thousand people. However, formula (5) assumes the use of a random sample of representatives of the population for testing. In practice, in Russia and in the world, mainly representatives of risk groups are tested (those who have been in contact with patients, who have arrived from disadvantaged regions with signs of SARS, tourists, medical workers, etc.), among whom the concentration of cases is naturally higher. This situation with testing explains the above discrepancy and suggests that to estimate the total number of infected people in the country, separate studies of the representative samples of the entire population are required.

¹⁰ A closer analysis shows that the rate of exponential growth in the number of cases during this period is mainly due to the exponential increase in the number of tests performed.

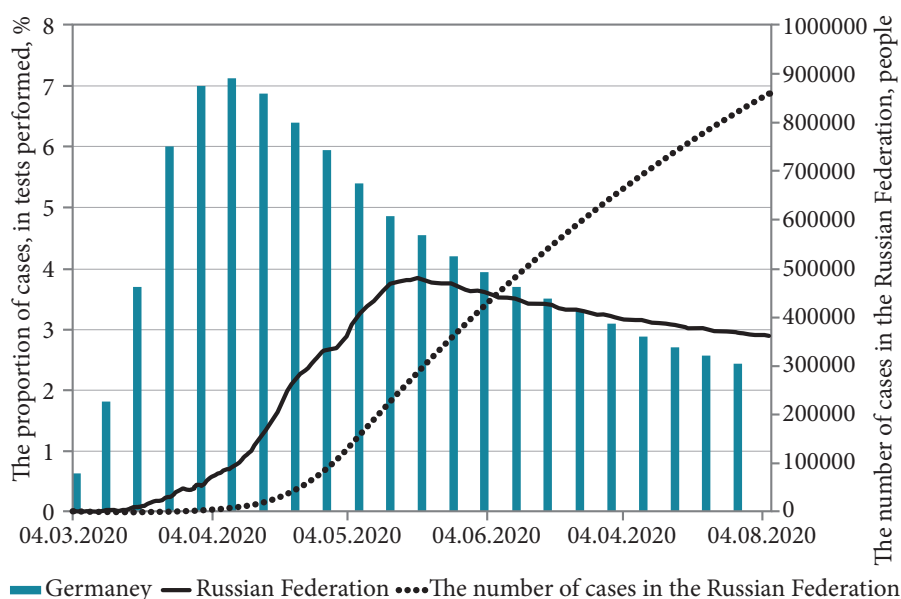


Figure 2. The proportion of detected COVID-19 diseases in the tests performed as of August 5, 2020 and the number of cases

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

Figure 3 shows the graph of the probability density distribution of the proportion of cases in tests performed across the countries of the world, which shows a large spread of values. The median value for the countries of the world is 3.13%, the average value is 4.84%.

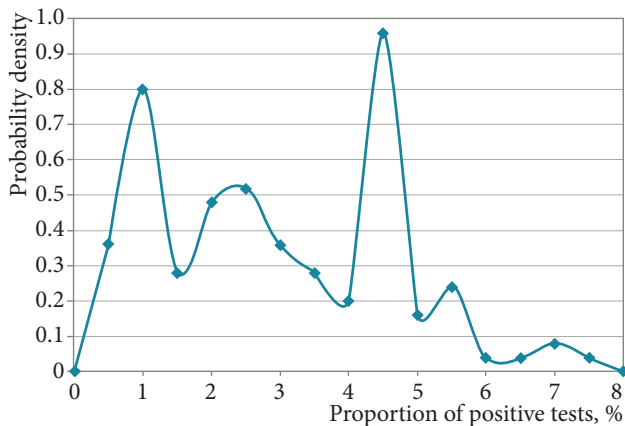


Figure 3. Probability density of distribution of the proportion of SARS-CoV-2 cases by country

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

Figure 4 shows the number of tests performed in Russia every day. Periodic fluctuations in Figure 4 reflect the activity of testing by days of the week, apparently related to the established COVID-19 testing procedures. Figure 4 shows that since May 30, the ratio T/N has stabilized at the average level of 0.194%, while the median value has remained practically the same – 0.198%.

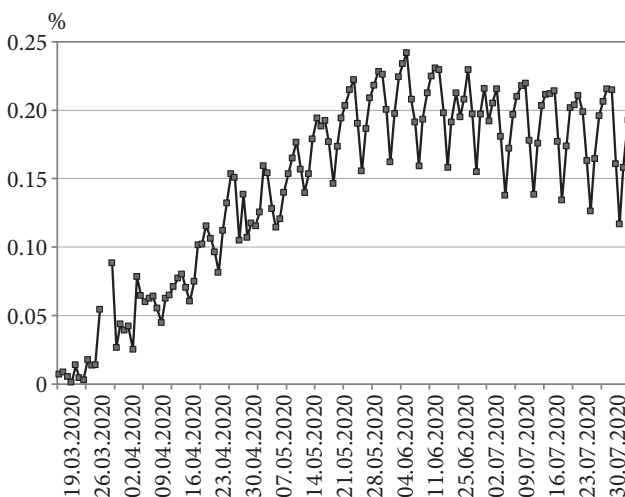


Figure 4. The ratio of daily tests and the population of the Russian Federation (%) $\frac{T}{N}$ – in (5).

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

To eliminate problems with the analysis of equations (4) in terms of the above-mentioned uncertainty of data on the number of infected individuals, various simplifying assumptions can be used:

1. The country's health care system detects all cases of the disease (more precisely, the proportion of those undiagnosed is negligible). In this case, group I is actually eliminated.

2. The ratio of the number of infected individuals and detected cases is approximately constant. For example, for Russia, according to the estimates above, the number of detected infected individuals is approximately 20–24% of the total number of infected people.

In what follows, we will focus on the second assumption. To accurately determine the ratio between the infected and detected active cases, it is required to solve equations (4), having previously determined the parameters by using, for example, regression models, analysis of medical statistics, etc. Relevant examples will be discussed below.

One more remark should be made. Following the works [5–9] cited above, a local temporal connection between all variables in equations (4) was assumed. In reality, this is not the case. For example, at time t , the individuals who were infected a certain number of days ago recover. Similarly, due to the long incubation period, at time t , the individuals who were infected up to 14 days ago are detected¹¹. Thus, equations (4) are retarded equations. The current size of the groups under consideration depends on their past size. To take this into account in equations (4), the following substitutions should be made¹²:

$$P_{IA}n_I(t) \rightarrow \int_0^t f_{IA}(t-\tau) \cdot \left[\frac{n_I(\tau)}{n_I(t)} \right] \cdot d\tau. \quad (7a)$$

$$P_{AD}n_A(t) \rightarrow \int_0^t f_{AD}(t-\tau) \cdot \left[\frac{n_A(\tau)}{n_A(t)} \right] \cdot d\tau. \quad (7b)$$

$$P_{AR}n_A(t) \rightarrow \int_0^t f_{AR}(t-\tau) \cdot \left[\frac{n_A(\tau)}{n_A(t)} \right] \cdot d\tau. \quad (7c)$$

$$P_{IU}n_I(t) \rightarrow \int_0^t f_{IU}(t-\tau) \cdot \left[\frac{n_I(\tau)}{n_I(t)} \right] \cdot d\tau. \quad (7d)$$

¹¹ <https://www.worldometers.info/coronavirus/> (date of access: August 10, 2020)

¹² For the discrete form of equations, the integration should be replaced by summation. The authors have kept the integral notation for the reasons of compactness. Thus, in their original form, Eqs. (4) are integro-differential retarded equations.

Here functions $f_{IA}(t - \tau)$, ..., $f_{IU}(t - \tau)$ specify (in the order of the equations): the density of the probability distribution of the incubation period (7a), the time from disease detection to death (7b), to recovery (7c) and (7d). The latter two will be determined by using the available statistics in section 3.2 below.

The probability distribution of the incubation period can be estimated by using the website data on the spread of coronavirus in the world¹³. Thus, it can be assumed, that function $f_{IA}(t - \tau)$ has the form shown in Figure 5 (obviously, the function must also include multiplier $P_{IA} \leq 1$, taking into account that some infected persons remain undetected).

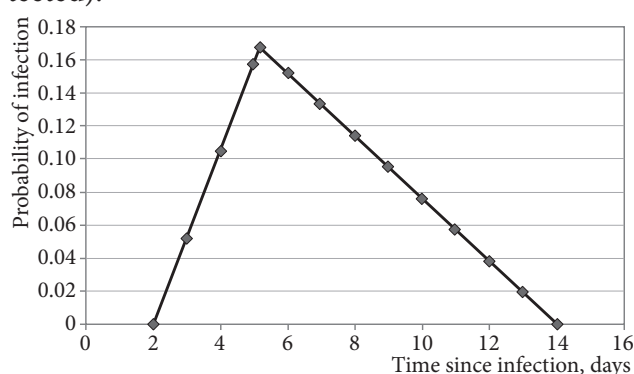


Figure 5. Probability of infection depending on the time elapsed from the moment of infection (incubation period). The function is shown without a constant multiplier for all points of P_{IA} (5)

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

Let us consider the process of transition of healthy individuals to the group of infected ones, which is the most important in the light of the pandemic. At the same time, further fate of the infected will not be considered (that is, transitions $I \rightarrow U$ are not analyzed). Then, from 4 (a) and (4b), one can find:

$$n_H(t+1) = n_H(t) - \frac{f \cdot n_I(t) \cdot n_H(t)}{n_H(t) + n_I(t)}. \quad (8a)$$

$$n_I(t+1) = n_I(t) - \int_0^t f_{IA}(t-\tau) n_I(\tau) d\tau + \frac{f \cdot n_I(t) \cdot n_H(t)}{n_H(t) + n_I(t)}. \quad (8b)$$

The latter terms in (8) appeared due to the decoding of probability P_{HI} (4a, b). In this case, an urn sampling scheme of two with returns is used as a model of contacts. The choice is made among

all the undetected infected and healthy and susceptible to disease individuals; f is the average frequency of contacts in the population, multiplied by the probability of infection by contact¹⁴. Thus, the presented formula assumes that an individual can become infected with a probability proportional to the fraction of the infected among all individuals with whom the said individual came in contact (Svirezhev, 1987).

We can attempt to determine the parameters (8b) on the basis of statistical data on COVID-19 cases. For this purpose, it should be taken into account that, despite the large absolute number of cases, their share in the population is small:

$$n_I \approx \frac{n_I}{n_H} \ll 1,$$

where (8) has the form:

$$\Delta n_H(t) = n_H(t+1) - n_H(t) = -f \cdot n_I(t) \quad (9a)$$

$$\Delta n_I(t) = -P_\alpha \left(d + \frac{T}{N} \right) \times \int_0^t f_{IA}(t-\tau) \cdot n_I(\tau) \cdot d\tau + f \cdot n_I(t). \quad (9b)$$

The above-mentioned multiplier P_{IA} (5) is clearly distinguished in formula (9b). Taking into account the above assumption that the ratio of the number of infected individuals and detected cases is constant, the construction of a regression model can be simplified. Namely, the statistical data on the number of detected cases provided by Yandex DataLens¹⁵ were used instead of the unavailable data on the number of the infected. The results are summarized in Table 1.

Table 1
Parameters of equation (9b) according to regression data

No.	Parameter	Value	Significance level of rejection of the null hypothesis according to t -statistics, %
1	$P_\alpha d$	0.12	9.1
2	$\frac{P_\alpha}{N}$	$4.45 \cdot 10^{-9}$	$5.5 \cdot 10^{-102}$
3	f	0.14	$5.5 \cdot 10^{-112}$

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>.

¹⁴ f can be decomposed by different social groups, regions, dates (for example, the frequency of contacts obviously increases on holidays), just as the frequency of contacts between teaching staff of schools and universities, doctors is higher than the average for the population due to the specifics of their work. However, this task is quite complex and requires special sociological and virological research.

¹⁵ <https://datalens.yandex.ru> (date of access: August 10, 2020)

¹³ <https://www.worldometers.info/coronavirus/> (date of access: August 10, 2020)

Using the data on the population of the Russian Federation¹⁶ and parameters from Table 1, P_α (reliability of disease detection – second line) can be determined: $P_\alpha = 65.3\%$. Then, from the first line, the proportion of individuals with the completed incubation period who independently consult doctors every day can be found: $d = 18.4\%$. Finally, $f = 0.14$. The comparison of the data calculated by model (9b) with the actual data is shown in Figure 6.

Finally, using equations (4)–(9), some of the usual characteristics of the epidemic can be expressed.

The spreading coefficient R_t :

$$R_t - 1 = \frac{\Delta n_I(t)}{n_I(t)} = f - \frac{P_\alpha \left(d + \frac{T}{N} \right)}{n_I(t)} \cdot \int_0^t f_{IA}(t - \tau) \cdot n_I(\tau) \cdot d\tau. \quad (10)$$

The condition $R_t < 1$, which is mandatory to remove the quarantine's restrictions, is reduced to the following equation:

$$n_I(t) \leq \lambda \cdot \int_0^t f_{IA}(t - \tau) \cdot n_I(\tau) \cdot d\tau, \quad (11)$$

where:

$$\lambda = \frac{P_\alpha \left(d + \frac{T}{N} \right)}{f}. \quad (11b)$$

The equal sign in (11) corresponds to a plateau – a growth termination of the number of the infected. Using the mean value theorem standard for estimation of the value of the integral in (11) (Fikhtenholts, 1970), one can find:

$$\lambda \geq \frac{n_i(t)}{SMA_T[n_I(t)]} \leq 1. \quad (12a)$$

Here $SMA_T[n_I(t)]$ is the simple moving average with a period equal to the incubation period. In accordance with (12a), the sufficient condition for reaching a plateau has the following form:

$$\frac{P_\alpha \left(d + \frac{T}{N} \right)}{f} \geq 1. \quad (12b)$$

The meaning of this condition is obvious, in a time unit (e.g. a day) the number of infected people identified and isolated should exceed the number of newly infected. We use the data from Table 1 and take into account that the most controllable parameter in (12b) is the number of tests carried out per time unit. It can be shown that to reach a plateau and subsequent transition to a decline in incidence, at least 4.6 million tests must be performed daily (provided that all other factors remain the same). Of course, the current testing

¹⁶ <https://rosstat.gov.ru/> (date of access: August 10, 2020)

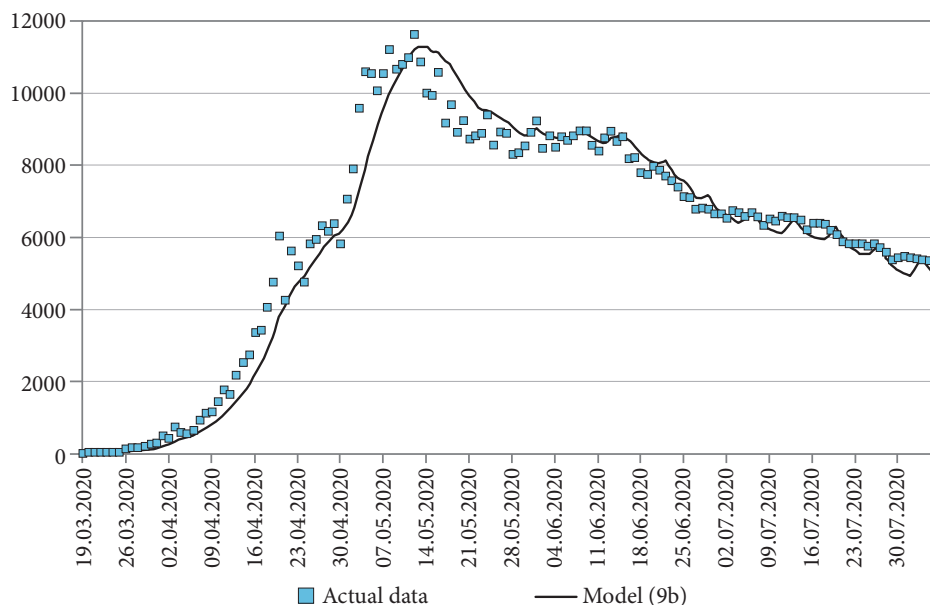


Figure 6. Comparison of the actual data on the identified infected persons and model (9b) with the parameters of Table 1. The coefficient of determination is 0.99

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

figures for the Russian Federation are much more modest. Given the cost of testing, the alternative is obvious – the transformation in the right direction of all the factors listed in the previous paragraph, including reduction in the number of contacts (social distancing) and the probability of infection upon contact (mask and glove mode, disinfection).

The exact solution to equations (4)–(9) is a rather complicated problem, which will not be considered in this work. However, as the above assessments show, even without such a solution, one can obtain useful information for analyzing the development of the COVID-19 pandemic in Russian regions.

In the following sections, we are going to discuss the dependence of the parameters of models (9b) and (4), determining the main characteristics of the pandemic on socio-economic parameters and characterizing Russian regions.

2. Methods of analysis

The following methods were used to analyze the entire set of collected data on the COVID-19 pandemic and the socio-economic situation in Russian regions.

We apply such methods of statistical analysis as correlation analysis of parameters characterizing morbidity and factors describing the socio-economic state of the regions and multivariate regression.

Correlation analysis was used to determine the degree of influence of regions' socio-economic characteristics on the epidemiological situation in these regions.

By using a multivariate regression, we were able to determine the parameters of the mathematical epidemic spread model proposed in Section 1 based on the comparison of the number of cases predicted by the model and the actual number of cases registered in Russia.

We also applied the following methods of data mining (Barsegyan et al., 2007; Leskovec et al., 2020):

- clustering by using self-organizing Kohonen maps (Debock, & Kohonen, 2001), which led us to divide all Russian regions into clusters – groups with similar characteristics of the incidence of COVID-19;

- hierarchical clustering (Zhambu, 1988), which enabled us to separate the “abnormal” clusters that differ significantly from the others in terms of morbidity characteristics and exclude

them from the sample for subsequent regression and correlation analysis.

The data from the website on coronavirus in the Russian Federation¹⁷, the service with Yandex DataLens¹⁸ datasets as well as the data from Rosstat for Russian regions (2018) were used for the analysis.

3. Results of the analysis of factors affecting the incidence of COVID in Russian regions

This section provides a comparative analysis of Russian regions in terms of the incidence of COVID-19 and socio-economic factors, which, in accordance with the theoretical concepts described in Section 2, can affect the characteristics of the pandemic.

3.1. Characteristics of the COVID-19 epidemic in Russian regions

The selection of adequate characteristics of the analyzed objects is a prerequisite for successful application of data mining methods. In the future, in addition to the obvious characteristics such as the infection rate (number of cases per 1,000 people) and the mortality rate (proportion of deaths among those infected), we are going to use the following indicators:

- as the growth rate of the infected (the popular parameter R_t) see (10);
- the growth rate of the recovered

$$i_R(t) = \frac{N_R(t+1)}{N_R(t)},$$

defined by equation (4f);

- the growth rate of lethal outcomes

$$i_D(t) = \frac{N_D(t+1)}{N_D(t)},$$

defined by equation (4d);

- the time from the moment the infected person is identified until the moment of his or her recovery or death.

Let us now consider these characteristics in more detail.

3.1.1. General characteristics of COVID-19 pandemic in Russian regions

Figure 7 shows the distribution of the mortality rate by Russian regions. For comparison, the

¹⁷ <https://стопкоронавирус.рф> (date of access: August 10, 2020)

¹⁸ <https://datalens.yandex.ru> (date of access: August 10, 2020)

figure shows a normal distribution with the same mean and standard deviation.

As can be seen, the distribution of the lethality level differs significantly from the normal one (this is also confirmed by the corresponding check using the Pearson, Kolmogorov-Smirnov criteria (Ivchenko, & Medvedev, 2014)).

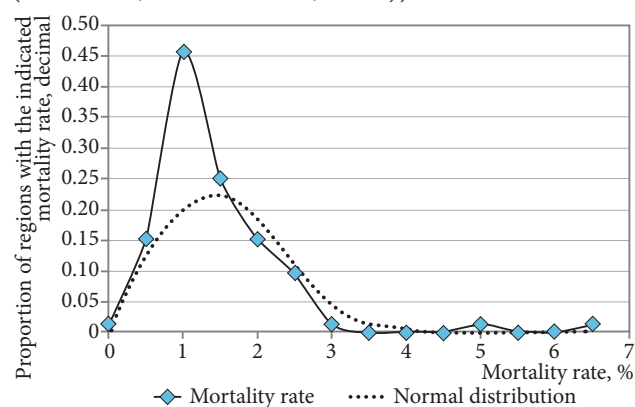


Figure 7. Distribution of the mortality rate by Russian regions

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

The level of infection is shown in Figure 8. For comparison, the normal distribution curve is shown by the dotted line.

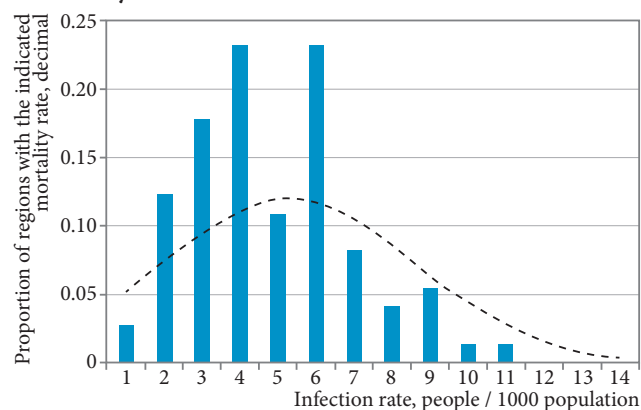


Figure 8. Distribution of the COVID-19 infection rate by Russian regions

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

The significant difference between the data presented in Figures 7 and 8 from the normal distribution indicates the presence of certain non-random factors that distinguish the course of the pandemic in various regions. This is evidenced by the data on the geometric mean growth indices of the number of the infected, the recovered, and lethal outcomes since the beginning of the pandemic in Russia. The corresponding data as of the end of July are presented in Figure 9.

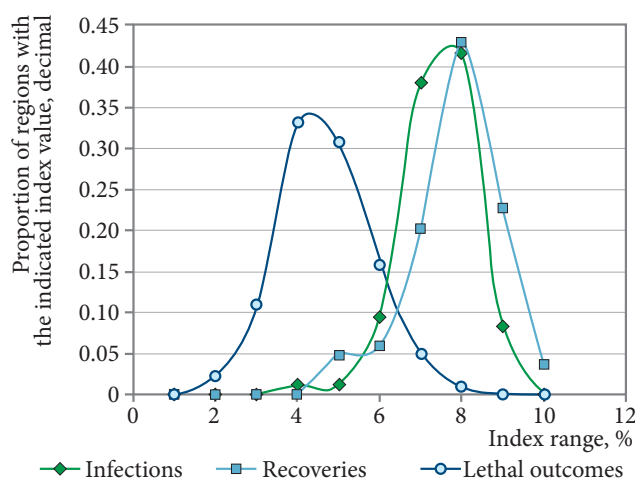


Figure 9. Growth rates of indices for cases: infections – $[R_t - 1]$, recoveries – $[i_R(t) - 1]$ and lethal outcomes – $[i_D(t) - 1]$

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

As can be seen, the spread of indices by regions is also quite noticeable. In this regard, it is of interest to analyze the factors affecting the rate of development of the pandemic. It can be concluded from the theoretical model in Section 2 that the presented indices are determined not only by the pathogenicity of the virus in a particular region but also by the socio-economic parameters of the latter, which affect the probabilities of transitions between different groups of pandemic participants in equation (9).

3.1.2. Analysis of the time of recovery and death of patients in Russia

In this section we are going to consider one more characteristic of the pathogenicity of SARS-CoV-2 – the duration of the disease. The initial data for the analysis of the time of recovery or a lethal outcome were obtained from Yandex DataLens¹⁹. The period from March 2, 2020 to July 18, 2020 was analyzed for the regions of Russia and from January 22, 2020 to July 17, 2020 for the world. Data processing was performed by using the scripts developed by the authors in the Python language in the Anaconda data analysis package.

According to the data mining standard CRISP-DM (Chapman et al., 2000), the obtained initial data required processing and preparation for analysis. Therefore, first, the countries in which there were no records of the recovered

¹⁹ <https://datalens.yandex.ru> (date of access: August 10, 2020)

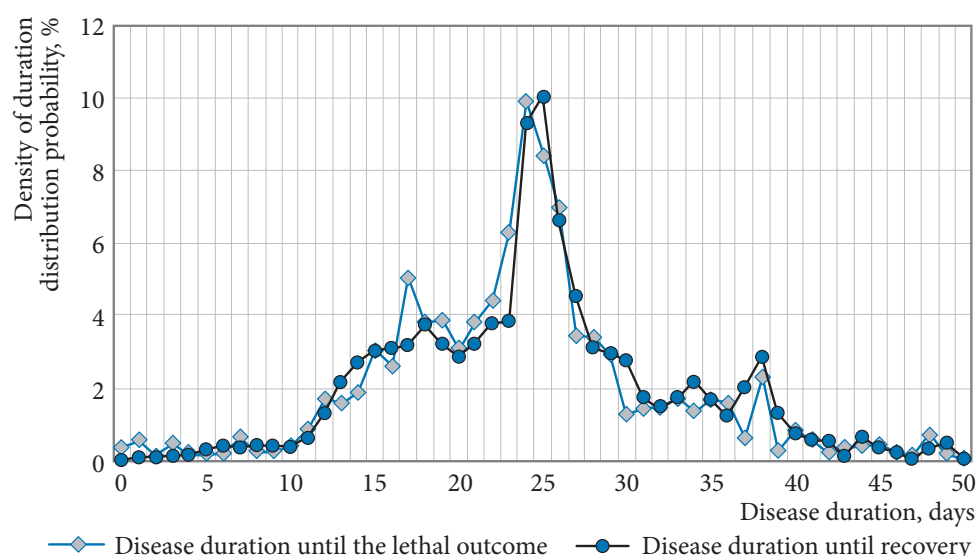


Figure 10. Distribution of disease duration by Russian regions

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

people²⁰ or the presented values differed significantly from the values for most countries²¹ were excluded from the data sample. Such distortions of statistics lead to an abnormally high mortality rate. We excluded regions and countries where the mortality rate exceeded 50% from the sample for analyzing the duration of the disease. Moreover, subjects with an unrepresentative number (less than 2,000) of closed cases – recoveries and lethal outcomes – were also excluded from the sample. Finally, the erroneous initial data were corrected, for example, the negative values of the numbers of new cases, recoveries, and lethal outcomes.

To determine the duration of the disease, the FIFO method (First In – First Out) – the first to enter the group of active cases is the first to leave it – recover or die, was used as the only available method according to official statistics. Thus, in each sample analyzed, the total number of recoveries and lethal outcomes was equal to the number of cases.

To determine which of the groups a closed case belongs to, the Monte Carlo procedure was applied, namely, the closed cases were divided into groups “Recovered” or “Dead” with the probability equal to the probability of recovery or death by the time point in question. The latter probability was determined by the ratio of the total number of

recoveries (lethal outcomes) to the total number of the infected (detected). The results obtained in the form of the distribution of the disease duration until recovery or death for Russian regions are shown in Figure 10.

For comparison, Figure 11 shows the data for the countries included in the sample formed in accordance with the above rules.

The presented figures show that the disease duration until recovery or death does not differ statistically significantly. On the contrary, the data on the disease duration for different countries and different Russian regions differ significantly. In particular, in Russia, the most probable disease duration is higher than in other countries (with a significantly lower mortality rate). Given the occasional foreign reports of the re-infections with COVID-19, there is some reason to believe that the existing treatment regimens for such patients may allow the completion of treatment prior to the patient's final recovery.

3.2. Clustering of Russian regions by disease characteristics

It should be noted that in different Russian regions, the disease can be caused by different strains of SARS-CoV-2²²; therefore, before analyzing the socio-economic factors influencing the

²⁰ For example, in Sweden, the number of cases is 77,281, the number of lethal outcomes is 5568, the number of recoveries is 0.

²¹ For example, in the UK, the number of cases is 294,803, the number of lethal outcomes is 42,477, the number of recoveries is 1312.

²² There have been reports in the media about the existence of various modifications of coronavirus in Russia (for example, https://radiokp.ru/obschestvo/skoltekh-nashel-v-ros-sii-9-shtammov-koronavirusa-kotorykh-net-v-drugikh-strakh_nid28655_au67au); however, at the time of this writing, it was not possible to find any serious scientific papers confirming this (quite probable) thesis.

spread of the disease, it is necessary to group the regions where the course of the disease has similar medical characteristics, that is, the subjects included in the cluster are closer to each other in these characteristics than to the subjects of other clusters. The mortality rate and the average time that elapses from the moment of infection to the recovery or death of the infected person were considered as such characteristics. The clustering of

Russian regions was carried out by using a self-organizing neural network (Debock, & Kohonen., 2001). As a result of data processing by the neural network²³ for all Russian regions included in the sample, the subjects were distributed across 10 clusters presented in Table 2. Table 2 also shows the significance levels of the variables for cluster

²³ The Deductor Studio software, academic version 5.3.0.88, was used for clustering.

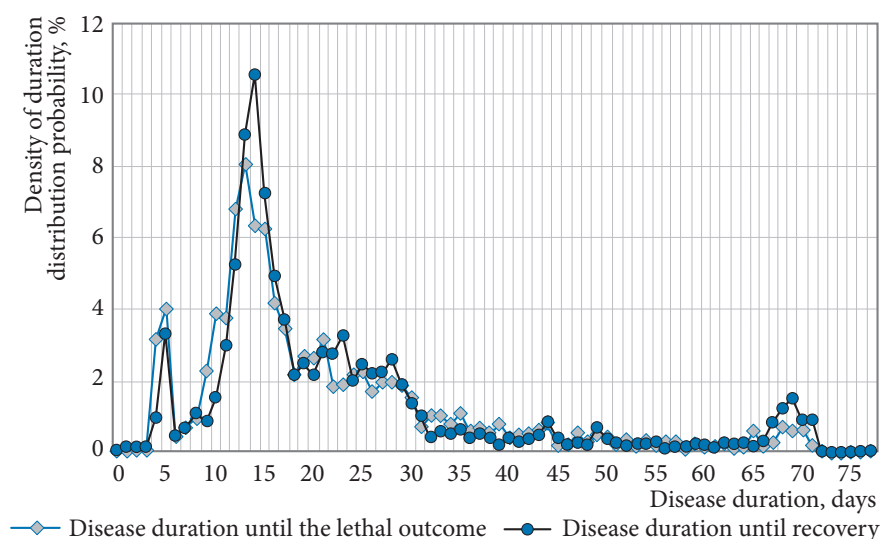


Figure 11. Distribution of disease duration by countries

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

Table 2

Results of the clustering of Russian regions by medical characteristics. The levels of significance of the variables for cluster formation are presented

Cluster number	Cluster members	Variable significance:		
		Time until recovery, %	Time until death, %	Mortality rate, %
0	Republic of Dagestan	76.6	95.3	100
1	Novosibirsk Region, Perm Region, the Republic of Ingushetia, Tver Region, Tula Region	83.9	85.5	97.9
2	Altai Territory, Belgorod Region, Bryansk Region, Voronezh Region, Zabaikalsky Region, Kirov Region, Krasnodar Region, Leningrad Region, Orenburg Region, Penza Region, Primorye Territory, Republic of Buryatia, Republic of Mari El, Republic of Sakha (Yakutia), the Republic of North Ossetia – Alania, Republic of Tatarstan, Republic of Khakassia, Khanty-Mansi Autonomous Area – Yugra	100	100	84.3
3	St. Petersburg	41.5	45.8	100
4	Vladimir Region, Kamchatka Territory, Krasnoyarsk Territory, Nizhny Novgorod Region, Smolensk Region	92.1	94.1	79.1
5	Moscow, Moscow Region	99.6	99.4	56.2
6	Karachayevo-Circassian Republic, Kursk Region, Murmansk Region, Komi Republic, Republic of Tuva, Ryazan Region, Tambov Region, Ulyanovsk Region, Khabarovsk Territory, Chuvash Republic, Yamal-Nenets Autonomous District	99.0	99.9	94.8
7	Arkhangelsk Region, Astrakhan Region, Ivanovo Region, Irkutsk Region, Kabardino-Balkarian Republic, Omsk Region, Oryol Region, Pskov Region, Rostov Region, Sverdlovsk Region, Stavropol Territory, Tyumen Region, Chelyabinsk Region	12.5	32.9	14.8
8	Volograd Region, Kaluga Region, Lipetsk Region, Republic of Mordovia, Samara Region, Saratov Region, Yaroslavl Region	96.1	75.2	84.0
9	Republic of Bashkortostan	84.3	83.5	66.8

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>.

formation (100% is the maximum significance). The higher the level of significance of a variable, the more likely it is that subjects with similar values of this variable will fall into the same cluster.

The clusters presented in Table 2 by the degree of similarity of the variables of the subjects included in them can be combined into a hierarchical structure shown in Figure 12.

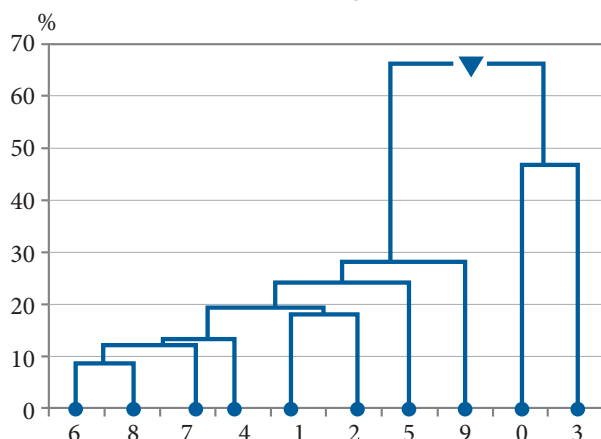


Figure 12. Hierarchical structure of the clusters uniting the subjects of the Russian Federation according to the medical characteristics of the COVID-19 disease

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>

The vertical axis represents the conditional distance (degree of difference) between the clusters (%). The less the given distance, the less the difference between the clusters. It is noteworthy that the clusters containing a small number of subjects (0, 3, 5, 9) are quite far from all the others. In the next section, when analyzing the dependence of the characteristics of the epidemic on the socio-economic factors of the region, we will focus on this hierarchy and carry it out only for subjects belonging to relatively close clusters.

4. Discussion. Correlation and regression analysis of socio-economic factors affecting the characteristics of the pandemic

As we pointed out at the end of the previous section, we will restrict ourselves to a correlation analysis of morbidity characteristics only for subjects included in clusters 1, 2, 4, 6, 7, 8. As can be concluded from equations (9)–(11), the rate of increase in the morbidity R_t is determined by the difference between the frequency of infection f and the frequency of detection and localization of patients:

$$R_t \approx 1 + f - P_\alpha \left(d + \frac{T}{N} \right). \quad (13)$$

At the same time, the very number of cases when the morbidity plateau is reached satisfies the equation:

$$0 \approx f \cdot (1 - n_I) \cdot n_I(t) - P_\alpha \left(d + \frac{T}{N} \right) \cdot n_I(t). \quad (14)$$

Hence:

$$n_I = \frac{f - P_\alpha \left(d + \frac{T}{N} \right)}{f}. \quad (15)$$

Thus, the morbidity rate R_t and the number of cases are proportional to the same value.

In accordance with the above, let us consider the factors able to affect parameters (15). At the same time, we will present below the results only in the case of a significant difference in the correlation coefficients and morbidity characteristics from zero. We will consider the correlation coefficients by groups of socio-economic factors. At the same time, since the values of the correlation coefficients are small in general and the distributions of the number of factors differ significantly from the normal, the quadrant correlation coefficient will also be calculated (Amosova et al., 2001). The factors for which the values of the standard and quadrant correlation coefficients differ significantly (and, moreover, have different signs) are discarded. All characteristics X_i for the sample of Russian regions before correlation analysis are reduced to the standard form:

$$\tilde{x} = \frac{(X - \bar{X})}{\sigma_X},$$

where \bar{X} is the average, and σ_X is the standard sample deviation. The results are presented below.

First, the presence of correlations of general characteristics of a region, such as square area, population, number of municipalities, population density, and road density (affecting the transport mobility of the population) with the characteristics of morbidity, was checked. Statistically significant correlations are presented in Table 3.

Thus, the risk group (high rate of development of the epidemic and a large number of cases) includes regions with a large population and a large number of municipalities.

Table 3

Statistically significant (with the confidence rate of more than 99% and the value greater than 0.3) correlation coefficients of the characteristics of the incidence of COVID-19 with the general characteristics of Russian regions, correlation coefficient/quadrant correlation coefficient

Characteristics of the disease	General characteristics of a region	
	Number of municipalities, units	Population size as of January 1, 2020, people
Number of cases	0.30/0.29	0.61/0.44
Infection rate	0.30/0.24	0.34/0.39

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>.

When analyzing the structure of the population, significant (with a confidence level more than 99%) correlations were found only for the

share of the urban and rural population (correlation coefficient is +0.3 for the share of the urban population and -0.3 for the rural population, the corresponding quadrant correlation coefficient is ± 0.27). The analysis confirms the above conclusion about the increased risk of disease for large settlements. This conclusion is confirmed by the analysis of the main economic indicators presented in Table 4.

The results of Table 4, in line with the previous data, indicate an increased risk of the pandemic in industrialized regions with large fixed assets, significant gross regional product, and developed retail trade. This is natural, since all of the above factors imply the concentration of a sufficiently large number of people in a relatively limited space (work premises, offices, shopping centers and shops), that is, an increase in the number of

Table 4

Significant (with the confidence rate of more than 99% and the value greater than the specified value) correlation coefficients of the characteristics of the incidence of COVID-19 with the main economic indicators of a Russian region, correlation coefficient/quadrant correlation coefficient

	Characteristics of the disease	
	Number of cases (with the value greater than 0.5)	Infection rate (with the value greater than 0.3)
Average annual number of employed, ths people	0.64/0.44	0.34/0.34
Gross regional product in 2017, mln rbs	0.64/ 0.39	0.36/0.34
Fixed assets in the economy (at full reported value; as of the end of the year), mln rbs	0.53/0.29	0.31/0.24
Volume of shipped domestic goods, works and services performed independently by types of economic activities*, mln rbs	0.6575/0.47	0.32/0.27
Retail turnover, mln rbs	0.63/0.49	0.31/0.34

* Average value by type: "Manufacturing", "Supply of electricity, gas and steam; air conditioning", "Water supply; wastewater disposal, waste collection and disposal, pollution elimination activities", "Retail trade turnover, mln rubles".

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>.

Table 5

Significant (with the confidence rate of more than 99% and the value greater than the specified value) correlation coefficients of the characteristics of the COVID-19 morbidity with other indicators of the activities in Russian regions, correlation coefficient/quadrant correlation coefficient

	Disease characteristics	
	Number of cases (with a value greater than 0.5)	Infection rate (with the value greater than 0.3)
number of employees of state bodies and local self-government bodies, people	0.64 /0.39	0.39/0.34
Number of unemployed, ths people	0.51/0.16	0.37/0.06(!)
Number of foreign citizens with a work permit, people	0.55/0.29	0.28/0.29
Number of enterprises operating in the field of culture, sports, leisure and entertainment, organizations, units	0.55/0.29	0.29/0.14
Number of students in educational programs of primary, basic, and secondary general education, ths people	0.58/0.34	0.34/0.29
Passengers travelling by railway, ths people	0.50 /0.39	0.25/0.29
Passengers travelling by public buses, mln people	0.61/0.44	0.29/0.44
Passenger turnover, public buses, mln pass./km	0.52/0.44	0.26/0.34

Source: the authors' calculations based on the dataset from Yandex DataLens service <https://datalens.yandex.ru>.

contacts – parameter f in (13–15). This feature of pandemics like that of COVID-19 indicates the significant risks that they can pose for economically developed regions. In this regard the risk management plan for business entities should be revised in order to take into account such threats.

When analyzing the impact of the labor market on the COVID-19 morbidity, the following significant correlations (with the confidence rate of more than 99%) are worth noting:

It is obvious that the factors listed above cover the main activities of modern urbanized territories, the state of the labor market, transport systems, and education. The lack of any significant correlation with the characteristics of health care and morbidity unrelated to COVID-19 is noteworthy. The use of such information makes it possible to determine all the parameters characterizing the development of the disease in equations (4)–(9) and thereby develop a consistent mathematical model corresponding to the actual data, whose numerical solution will enable us to consider various probable scenarios for the development of the pandemic in Russia. This task will be addressed in our future works.

Conclusion

The mathematical model proposed in this article allows for a logically consistent description of the development of the COVID-19 pandemic in Russia. It is noteworthy that the model relies on the data that can be obtained by using the national morbidity rate system and does not require the analysis of individual case histories, the data which are much more difficult to obtain. At the same time, an additional result was a check of the completeness and reliability of statistical data. Thus, when calculating the distribution of disease durations in the regions, the correspondence of the daily flows of the infected and the closed cases was checked. While in Russia only the isolated cases of inconsistency were observed, then the analysis of the data around

the world revealed quite a number of countries giving reasons to doubt the adequacy of the statistical recording of morbidity. It should be noted that the need and importance of organizing an adequate and operational statistical accounting system necessary for making management decisions during pandemics like COVID-19 can be considered as one of its lessons.

Due to the relevance of this issue, we decided not to present in detail the results and conclusions obtained in this article, leaving them for individual research.

The correlation analysis of the influence of socio-economic factors on the development of the disease indicates that COVID-19 and other similar pandemics are a serious challenge for modern civilization. Like the development of a virus in the cells of an organism that uses the normal processes to function, the pandemic develops by using the socio-economic processes that have developed and are necessary for a civilized society. In this regard, a kind of “vaccination” of socio-economic systems is required to reduce the rate of infection, the number of cases, or both indicators at the same time. This refers to the rational organization of transport services, cultural events, jobs, educational processes, and the entire socio-economic structure of the region. Certain experience in connection with quarantine measures has already been accumulated. However, to ensure the resilience of socio-economic systems to such diseases, evidence-based measures are needed that have a selective effect on the key risk factors for the development of the pandemic and take into account their long-term consequences. It could be recommended to analyze the developed strategies for the socio-economic development of the regions for the period up to 2035 from this point of view and make the necessary adjustments. Such activities and adjustments to strategies should be based on models such as those outlined in this paper.

References

- Abdo, M.S., Shah, K., Wahash, H.A., & Panchal, S.K. (2020). On a comprehensive model of the novel coronavirus (COVID-19) under Mittag-Leffler derivative. *Chaos, Solitons & Fractals*, Vol. 135:109867, doi: [10.1016/j.chaos.2020.109867](https://doi.org/10.1016/j.chaos.2020.109867)
- Amosova, N.N., Kuklin, B.A., Makarova, S.B., & Maksimov, Yu.D. (2001). *Probabilistic sections of mathematics*. St. Petersburg: Ivan Fedorov, 592.
- Bailey, N.T.J. (1975). *The Mathematical Theory of Infectious Diseases and Its Application*. Griffin, London.

- Barlow, N.S., & Weinstein, S.J. (2020). Accurate closed-form solution of the SIR epidemic model. *Physica D: Nonlinear Phenomena*, Vol. 408: 132540, doi: [10.1016/j.physd.2020.132540](https://doi.org/10.1016/j.physd.2020.132540)
- Barsegyan, A.A., Kupriyanov, M.S., Stepanenko, V.V., & Kholod, I.I. (2007). *Data analysis technologies. Data mining, Visual mining, Text mining, OLAP*. 2nd ed. St. Petersburg: BHV-Petersburg, 384.
- Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons & Fractals*, Vol.135: 109850, doi: [10.1016/j.chaos.2020.109850](https://doi.org/10.1016/j.chaos.2020.109850)
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide: [ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf](http://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf)
- Debock, G., & Kohonen, T. (2001). *Analyze financial data using self-organizing maps*. Moscow: Alpina, 317.
- Driessche, P., & Watmough, J. (2020). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, Vol. 180, 29–48, doi: [10.1016/S0025-5564\(02\)00108-6](https://doi.org/10.1016/S0025-5564(02)00108-6)
- Feller, W. (1964). *An introduction to probability theory and its applications*. Volume 1. Moscow: Mir, 498.
- Fikhtengolts, G.M. (1970). *Differential and integral calculus course*. Volume II. Moscow: FML, 800.
- Ivchenko, G.I., & Medvedev, Yu.I. *Mathematical Statistics* (2014). Moscow: LIBROKOM, 352.
- Leskovec, J., Rajaraman, A., & Ullman, J.D. (2020). *Mining of Massive Datasets*. Cambridge University Press, 3 ed., 565.
- Ndairou, F., Area, I., Nieto, J.J., & Torres, D.F.M. (2020). Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons & Fractals*, Vol. 135: 109846, doi: [10.1016/j.chaos.2020.109846](https://doi.org/10.1016/j.chaos.2020.109846)
- Rvachev, L.A. (1971). Experiment on machine forecasting of influenza epidemic. *Reports of the USSR Academy of Sciences*, volume 198, number 1, 68–70.
- Sinitsyn, E.V., Astratova, G.V., & Frishberg, L.A. (2011). Using information models to assess intellectual property. *Practical Marketing*, No. 5 (171), 28–35.
- Svirezhev, Yu.M. (1987). *Nonlinear waves, dissipative structures and catastrophes in ecology*. Moscow: Nauka, 368.
- Tolmachev, A. V., Sinitsyn, E. V., & Brusyanin, D. A. (2019). Transport system modelling based on analogies between road networks and electrical circuits. *R-economy*, 5(2), 92–98. doi: [10.15826/recon.2019.5.2.010](https://doi.org/10.15826/recon.2019.5.2.010)
- Whittle, P. (1955). The outcome of a stochastic epidemic – a note on Bailey's paper. *Biometrika*, Vol. 42, Issue 1–2, 116–122, doi: [10.1093/biomet/42.1-2.116](https://doi.org/10.1093/biomet/42.1-2.116)
- Xiaolei Zhang, & Renjun Ma and Lin Wang (2020). Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. *Chaos, Solitons & Fractals*, Vol. 135:109829, doi: [10.1016/j.chaos.2020.109829](https://doi.org/10.1016/j.chaos.2020.109829)
- Zhambyu, M. (1988). *Hierarchical Cluster Analysis and Compliance*. Moscow: Finance and Statistics, 345.

Information about the authors

Evgeny V. Sinitsyn – Doctor of sciences, Professor, Head of Department of Management, Graduate School of Economics and Management, Ural Federal University (19 Mira St., 620002, Ekaterinburg, Russia); e-mail: sinitsyn_ev@mail.ru

Alexander V. Tolmachev – CEO, Datatel-Ural (12/8 Sibirsky trakt, 620100, Ekaterinburg, Russia); Senior lecturer, Department of management, Graduate School of Economics and Management, Ural Federal University; e-mail: at@idtu.ru

Alexander S. Ovchinnikov – Doctor of sciences, Professor, Department of Theoretical and Mathematical Physics, Institute of Natural Sciences and Mathematics, Ural Federal University (19 Mira St., 620002, Ekaterinburg, Russia); e-mail: alexander.ovchinnikov@urfu.ru

ARTICLE INFO: received July 25, 2020; accepted September 7, 2020

Информация об авторах

Синицын Евгений Валентинович – доктор физико-математических наук, профессор, заведующий кафедрой менеджмента, Институт экономики и управления, Уральский федеральный университет (620002, Россия, г. Екатеринбург, ул. Мира, 19); e-mail: sinitsyn_ev@mail.ru

Толмачев Александр Владимирович – генеральный директор Датател-Урал (620100, Россия, г. Екатеринбург, Сибирский тракт, 12/8); старший преподаватель кафедры менеджмента, Институт экономики и управления, Уральский федеральный университет (620002, Россия, г. Екатеринбург, ул. Мира, 19); e-mail: at@idtu.ru

Овчинников Александр Сергеевич – доктор физико-математических наук, профессор, кафедра теоретической и математической физики, Институт естественных наук и математики, Уральский федеральный университет (620002, Россия, г. Екатеринбург, ул. Мира, 19); e-mail: alexander.ovchinnikov@urfu.ru

ИНФОРМАЦИЯ О СТАТЬЕ: дата поступления 25 июля 2020 г.; дата принятия к печати 7 сентября 2020 г.